

Семантический метапланировщик GRID

Барабаш Михаил¹, Поспешный Александр¹

¹ *Национальный Технический Университет Украины «Киевский Политехнический Институт»
пр. Победы 37, Киев, Украина, 03056*

Barabash.michael@gmail.com, pospishniy@gmail.com

Аннотация. В работе описаны базовые принципы построения гетерогенных сетей, а также проблемы эффективного, распределения и планирования ресурсов, для решения пользовательских задач. Предлагается создание перспективного семантического метапланировщика и приводится принцип его работы.

Ключевые слова

Метапланировщик, GRID, RDF, онтология.

Гетерогенные вычислительные сети нашли свое применение в решении широкого спектра научных задач, ряда национальных проектов в таких областях как прогнозирование природных и техногенных катастроф, медицинские исследования и распределенные хранилища данных. Частные корпорации также активно строят часть своих бизнес процессов с использованием гетерогенных вычислительных сетей, что дает право уверенно сказать что всего лишь чем через десять лет GRID технологии нашли свое применение практически в всех областях. Основной задачей GRID является объединение гетерогенных вычислительных ресурсов в единую сеть с возможностью предоставления прозрачного и безопасного доступа к ним при решении научных, национальных и бизнес задач.

Для высокоуровневого описания всех доступных ресурсов GRID используются LDAP сервера объединенные в иерархическое дерево, корнем которого является информационный сервер самого верхнего уровня, к которому можно подключиться и выполнить запрос. Для обмена файлами между ресурсами GRID, используется модифицированный FTP протокол - GRIDFTP. Для обеспечения авторизованного доступа к разделяемым ресурсам, чаще всего пользователи объединяются в так называемые виртуальные организации, где им выдаются x509 сертификаты, на основании которых службы GRID предоставляют запрашиваемые при решении задач ресурсы.

Практически со своего зарождения, GRID сообщество активно ведет работу по стандартизации и унификации базовых технологий используемых при проектировании промежуточного ПО, при помощи которого строится GRID. Однако этот процесс идет не так быстро, как нам бы всем хотелось, да и так сложилось исторически, что у каждого региона есть свои предпочтения при выборе и проектировании промежуточного ПО, на котором в дальнейшем строятся национальные GRID проекты.

Основной проблемой распределения доступных ресурсов GRID сети, является нахождение и предоставление наиболее подходящих вычислительных узлов для решения разнообразнейших пользовательских задач, которые как правило требуют специфичное для решаемой задачи дополнительное программное обеспечение, начиная от версии компилятора и заканчивая предустановленными специализированными библиотеками и САПР приложениями и все это на различных программных и аппаратных платформах.

Практически любая GRID система выполняет задачу в следующей последовательности. После попадания задачи в **очередь GRID сегмента**, выполняется анализ ее требований и дальнейший подбор наиболее подходящего вычислительного элемента. Для этого **метапланировщик** выполняет LDAP запрос к **службе мониторинга** и в случае нахождения подходящего кандидата, эта задача попадает в **локальную очередь планировщика** вычислительно узла, где и ожидает начала своего выполнения.

Существующие метапланировщики такие как WMS, GridWay входящие в состав gLite и Globus [1,2], работают в паре с информационными супермаркетами, в которых хранятся закешированные данные о текущем состоянии системы в обслуживаемом метапланировщиком GRID сегменте. Информационный супермаркет

строится, как правило, на реляционной базе данных и процесс подбора ресурсов сводится к обычному сопоставлению требований задачи и свойств того или иного гомогенного кластера, при этом в LDAP, хранится информация, описывающая лишь базовые характеристика кластера (общее количество ядер, сколько из них свободно, состояние локальной очереди, архитектура и тип процессоров). На рисунке 1 ниже, показана базовая схема взаимодействия основных компонентов GRID. Для упрощения на рисунке не показаны сервисы обслуживающие виртуальные организации и службы управления репликами данных, а внимание уделено именно процессу планирования.

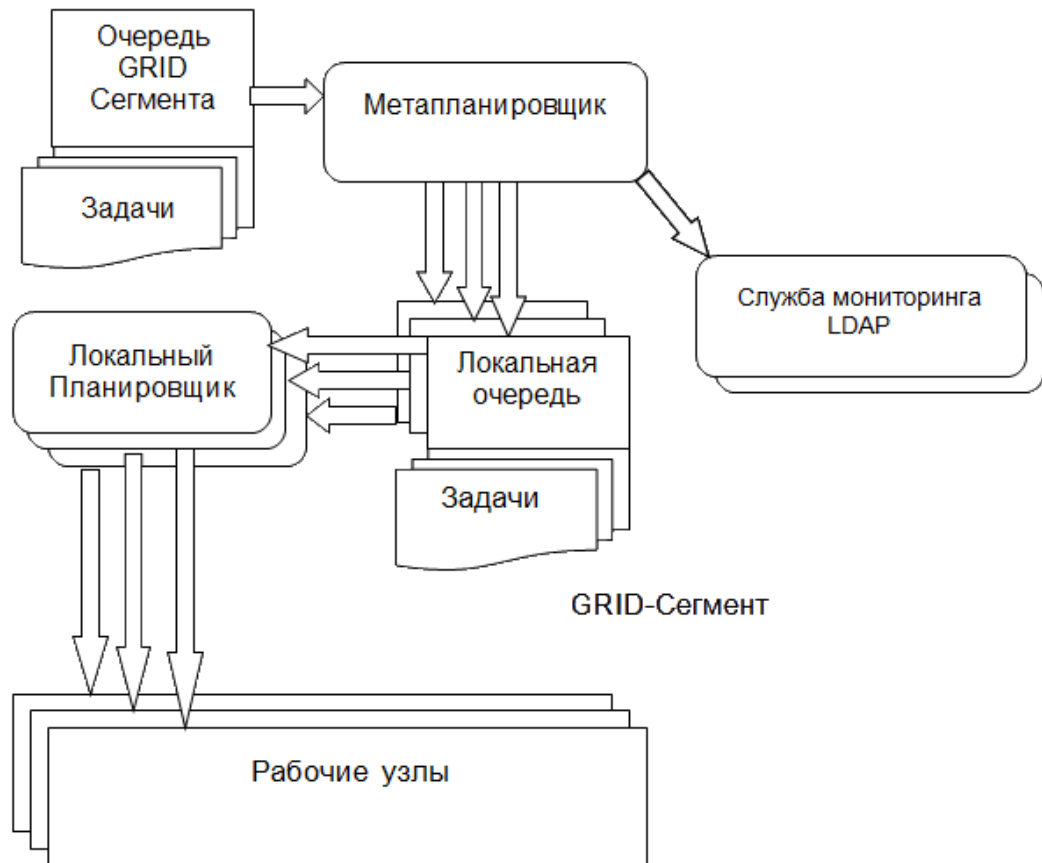


Рис. 1. Схема взаимодействия компонентов GRID

Как было сказано выше, пользовательские задачи могут быть из самых разных доменных областей, и чтобы метапланировщик был способен эффективно подбирать под них ресурсы, ему и самому необходимо обладать соответствующими доменными знаниями. А так как GRID должен предоставлять как можно более широкий спектр решаемых задач, то метапланировщику необходимо быть в состоянии пополнять самостоятельно свои знания о предметных областях задач пользователей, которых он обслуживает.

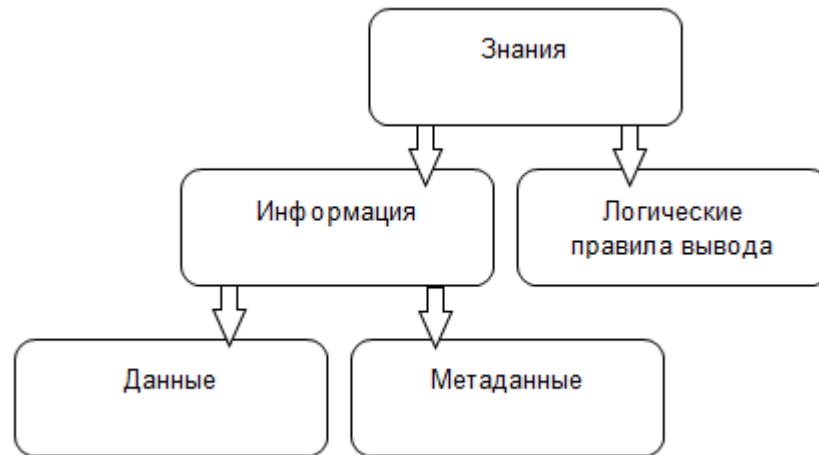


Рис. 2. Знання, которыми должен обладать метапланировщик

Знання - это информация и правила вывода, в то время как информация – это данные и метаданные, а метаданные – это данные о данных.

Для формализации области знаний используются онтологии, состоящие из специальных структур данных, описывающие все релевантные классы объектов, а также теоремы, ограничения и связи, между ними. Одной из первых практических попыток формализации знаний в сети, была инициатива создания семантической паутины *Semantic Web*[3]. Хотя этот проект еще не завершен, а возможно и не может быть завершен в настоящее время ввиду своей глобальности и существующих технических ограничений, он показал что возможно впринципе описать доменную область знаний в доступном для компьютера виде и в последствии применять эту онтологию для дальнейшего логического вывода дополнения существующей базы знаний. При этом были предложены инструменты, языки и форматы способные описывать различные области знаний. Для описания метаданных используется Resource Description Framework (RDF)[4], представляющий утверждения о ресурсах в удобном для компьютера виде.

Таким образом, наделенный знаниями метапланировщик сможет самостоятельно из одной информации выводить новую. Однако, для того чтобы это было возможно, необходимо реализовать доступ метапланировщика к RDF модели вычислительной GRID сети. Построение такой модели позволит решить ряд таких стратегических на сегодняшний день задач, как:

- Легкость обмена данными между различными службами GRID
- Гибкость модели данных

И позволит создать принципиально новый метапланировщик, который будет способен:

- Производить семантический поиск ресурсов для каждой пользовательской задачи
- Делать логический вывод новых фактов о требованиях задач к ресурсам

Таким образом, предлагается создать в паре две совершенно новые GRID компоненты: это **семантический информационный сервис**, а также **семантический метапланировщик**, который будет его использовать. Семантический метапланировщик будет анализировать входящие пользовательские задачи, строить по их требованиям запросы к семантическому информационному сервису и при необходимости расширять существующую доменную область знаний. Предполагается, что после подбора потенциальных ресурсов, удовлетворяющих требованиям задачи, метапланировщик выполнит отдельно запросы к локальным планировщикам отобранных ресурсов и расширит существующую онтологию такими данными как: аппроксимированное время ожидания и выполнения задачи, время затрачиваемое на передачу данных, всевозможные корреляции с уже запущенными задачами. Такая пара сервисов будет способна взаимодействовать с любыми GRID сегментами, путем создания специфичного для каждого из них провайдера. На рисунке 3 показывается способ взаимодействия метапланировщика с семантическим информационным сервисом. Данный подход позволит более эффективно распределять существующие ресурсы GRID, а также производить общее моделирование системы.



Рис. 3. GRID сегмент управляемый семантическим
метапланировщиком

Список литературы

- [1] gLite, <http://glite.web.cern.ch/glite>
- [2] GridWay Metascheduler, <http://www.gridway.org>
- [3] Semantic web, <http://semanticweb.org>
- [4] Resource Description Framework <http://www.w3.org/RDF/>